

基于伸展树的无线传感器网络事件区域检测

张书奎^{1,2}, 崔志明^{1,2}, 樊建席¹, 张文哲¹

(1. 苏州大学计算机科学与技术学院, 江苏苏州 215006;

2. 江苏省现代企业信息化应用支撑软件工程技术研发中心, 江苏苏州 215104)

摘要: 本文提出一种基于伸展树的事件区域容错检测算法, 该算法在构成的分布式融合树上实施, 由多元回归方程对事件发生的可能性进行评估, 利用被检测事件的时空相关性, 使得在缺少传感节点的区域, Sink 节点也可以获得事件发生的状态信息. 理论分析及仿真结果表明, 即使节点发生故障的概率较高, 提出的检测算法也能获得较高的精度和较小的误判率, 表明在事件区域容错检测中应用该方法是可行的.

关键词: 无线传感器网络; 伸展树; 容错检测; 多项式回归

中图分类号: TP393.02 **文献标识码:** A **文章编号:** 0372-2112 (2010) 2A-194-07

An Event Region Detection Scheme Based on Splay Tree in Wireless Sensor Networks

ZHANG Shu-kui^{1,2}, CUI Zhi-ming^{1,2}, FAN Jian-xi¹, ZHANG Wen-zhe¹

(1. School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China;

2. Jiangsu Province Support Software Engineering R&D Center for Modern Information Technology Application in Enterprise, Suzhou, Jiangsu 215104, China)

Abstract: This paper investigates a kind of algorithm called splay tree based event region fault-tolerant detection algorithm (STERD) for Wireless Sensor Networks (WSNs). Since the algorithm is implemented over a splay tree in a distributed manner, a polynomial regression addressed these problems of event region detection by having a splay tree of sensor nodes. The sink can get attribute values in regions devoid of sensor nodes for attribute values that show smooth spatio-temporal correlation. Theoretical analysis and simulation results indicate that the proposed STERD can attractively obtain the high event region detection accuracy and considerably reduce the false alarm probability even when fault nodes exist.

Key words: wireless sensor network; splay tree; Event detection; polynomial regression

1 引言

无线传感器网络(WSN)的一项基本任务是检测和报告发生在特定区域内的各种感兴趣事件. 某事件的发生认为是环境状态(如温度、湿度、压力等)的异常改变, 它可能以多种方式出现. 随着时间的推移, 若传感器读数保持平稳, 则认为此环境对传感器检测来说是时空关联的^[1], 可见, 位于同一区域中邻近传感器的读数关系密切. 当某时段传感器感知的读数偏离了正常值, 或者邻居节点的读数大大超过了预定义的阈值, 则可能是某事件发生或是传感器产生了故障^[2,3]. 但是, 传感器节点报告错误读数的原因是多方面的(如与邻居传感器的读数不同或虽然没有超过预先设定的阈值但却与它在上一时间间隔感知读数不同等). 若发生了通信或硬件

故障, 这些错误会导致节点无效. 意外破坏或恶意通信线路的改变也会导致无效节点的产生. 环境受其它因素的影响, 传感器也可能产生瞬时的错误读数. 特别是部署在粗糙环境中的动态传感器网络进行感兴趣事件的检测, 由于网络拓扑结构发生变化, 也会导致错误读数的产生. 所以, 在 WSN 中, 保证原始数据的可信性, 消除错误读数的影响, 是事件区域检测需要解决的关键问题之一.

文献[4]提出了事件区域检测的一种分布式局部算法. 该算法假设事件是空间相关而错误是空间不相关的, 且每个传感器发生错误的概率相同. 作为一种局部算法, 它只需要每个传感器与自己的邻居交换读数, 从而获得邻居节点检测事件发生的概率, 利用 Bayesian 分析, 判定发生的是故障还是事件. 文献[5]则改进了文

收稿日期: 2009-07-10; 修回日期: 2010-02-08

基金项目: 国家自然科学基金(No. 60873047); 江苏省自然科学基金(No. BK2008154, No. BK2009116); 江苏省现代企业信息化应用支撑软件工程技术研发中心开放基金(No. SX200903); 2009年江苏省普通高校研究生科研创新计划(No. CX09B_025Z); 苏州大学科研预研基金项目

献[4]的算法,考虑了传感器误差和错误引起的对事件判断的影响以及如何选择合适的邻居数目,达到容错的同时减少数据交换的目的.文献[4~8]都是基于概率分析的方法,需要假设每个传感器发生错误的概率相同,而且需要传感器进行复杂的运算.文献[9]提出了一种相对简单的非概率的检测算法,通过两轮投票以及少数服从多数原则(majority-voting),确定发生的是故障还是事件.关于事件边界检测的研究,文献[10,11]认为,事件的边界是指事件区域与正常区域的分界线.如果一个传感器的邻居中既有异常又有正常的读数,那么就认为该传感器位于事件区域的边界.还有一类工作关注于错误诊断^[12],这类工作关注的是如何确定错误发生的范围或者引发错误的错误源,而不是事件发生的范围.文献[13]介绍了一种基于簇头构建4叉树的方法进行边界检测,每片叶子与长方形区域相对应.文献[14,15]是使用一种基于分类的边缘检测机制,由线性多项式描述事件边界,由此判定事件是在检测区域的里面还是外面.然而,一个线性多项式不可能获取事件发生的精确细节.与前述方法不同,本文采取的是一种易于构建和扩展的二叉查询树,该查询树可以覆盖整个感兴趣的区域,每个感知节点仅报告自己的感知数据到最邻近的树节点,能快速将事件发生的信息从感知节点传输到 Sink.通过构建多元回归模型,以事件的时空关联特性检测整个区域,获得事件区域检测的估计值.

2 树的构建

伸展树的构建是本文所提出事件区域检测过程的基础.其目标是减少数据传输量,提高检测精度.对数据融合算法的理论分析表明^[16],在完全融合的情况下,寻找最优融合树的问题等同于求解最小 Steiner 树的 NP 完全问题.为此我们需要在计算能耗与传输能耗之间进行权衡.下面讨论融合树的构建过程.

2.1 网络模型

设 N 个资源受限的静态传感器节点随机地部署在检测区域 $R = (r \times r)$ 内,用集合 $S = (s_1, s_2, \dots, s_n)$ 描述,其中 s_i 表示第 i 个传感器,如图 1 所示.每个节点通过三角剖分^[14]都有它的位置信息,节点 s_i 位置用 (x_i, y_i) 表示,且每个节点具有唯一的 ID,相同的计算通信能力和能量资源.节点通过时间同步服务^[15]达到宽松的时间同步,通信接入采用 CSMA/CA 以减小信道冲突.本文的目标是在此 N 个节点的网络中构建融合树(Aggregation Tree, AT),AT 由 N_t 个节点组成,叫做树节点,树节点用于接收并融合数据,其余 $(N - N_t)$ 个节点叫做非树节点(Not tree node, NT),每个 NT 节点感知指

定区域事件属性的变化,并将感知数据传输到它最邻近的树节点.构建的 AT 扩展到整个网络,以便 N_t 个树节点均匀一致地分布在网络中,这样可以确保检测属性值以尽可能小的跳数由 NT 节点传送到对应的树节点,从而延长网络的寿命.为简便起见, P_{event} (在图 1 中矩形框里的虚线表示)代表事件, R_{event} 表示事件区域且 $R_{event} \subseteq R$.正常情况下,AT 覆盖检测区域里所有的事件. R' 定义为 R 中没有被覆盖的部分,即 $R' = R - R_{event}$.

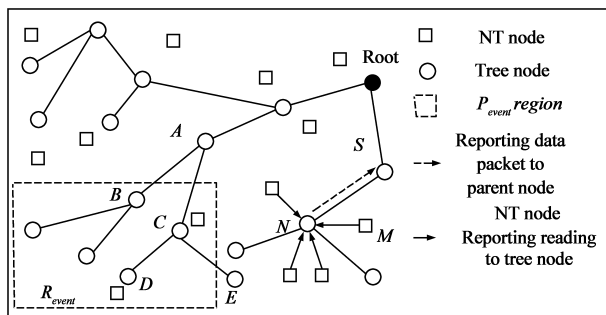


图1 网络模型

2.2 树的生成

某事件的发生会触发网络中的部分节点使其读数异常,这可能是一个孤立点,也可能是多个.为确保 AT 扩散到整个网络,感知值由传感节点通过较小的跳数传输到相应的树节点,尽可能保持分散节点的拓扑稳定性以维持原有较好的感知覆盖范围,为此引入了 Voronoi 图以及相关 Delaunay 三角网络来描述感知网络拓扑^[17],并基于 Delaunay 三角定义,构建以事件中心节点为根的感知网络融合树-伸展树.伸展树是一种二叉树,其优点在于不需要记录用于平衡树的冗余信息.设 e 为平面上的点,则

$$VR(e) = \{p \in R^n \mid d(p, e) \leq d(p, e'), \forall e' \neq e, e' \in E\} \quad (1)$$

称为 Voronoi 多边形.则 Voronoi 图定义为

$$VD(E) = \bigcup_i VR(e_i) \quad (2)$$

即平面上所有 Voronoi 多边形的集合,而 Delaunay 三角网络则为连接所有相邻的 V -多边形的生长中心所形成.Delaunay 三角网具有很多重要的性质^[17],通过 Delaunay 表达不仅可以获得每个节点的邻节点信息,而且可以用来查找最接近的节点.基于传感器网络的 Delaunay 描述,构建以事件中心节点为根的伸展树.设目标区域为 A ,区域中的感知节点集为

$$S = \{s_i(x_i, y_i) \mid s_i \in A\} \quad (3)$$

其中, (x_i, y_i) 为节点 s_i 的位置坐标.由节点集 S 构成的网络对应的加权无向图为 G ,各边所对应的权值为节点间距离.并设感知区域外存在点集 $K = \{k_i(x_i, y_i) \mid k_i \notin A\}$,则目标区域中以节点 s_i 为中心相对于点集 K 的节点伸展树定义为 T ,有

$$T(s_i - > K) = \bigcup_i \text{path}(s_i \rightarrow k_i), k_i \in K \quad (4)$$

其中, $\text{path}(s_i \rightarrow k_i)$ 为无向图 G 中从 s_i 到 k_i 的节点间的最大跨度路径, 其长度为 l . 在此路径上, 各节点间的最小距离大于或等于 G 中任何其它从 s_i 到 k_i 路径上节点间的最小距离. 最大跨度路径反映了两节点之间具有伸展性的一条通路. 需要指出的是, 在一个特定的无向图 G 中, 两点间的最大跨度路径并不惟一, 可能存在多条. 而对于不同的域外节点集, 相应的以根节点为中心的伸展树也并不惟一.

设树的深度为 p , 树节点存储感知的属性值, 这样的树被认为是平衡的, 它减少了数据丢失增加了数据融合的精确性^[15]. 算法 Form_AT 是由给定的深度创建伸展树. 每当一个节点要选择它的两个孩子时, 就选择跨度最大的两个节点, 可确保树在扩散时覆盖尽可能多的感兴趣区域, 伸展树形成后, 各子区域中所有剩余节点向距离自己最近的树节点发送数据. 本文通过三种消息 Beacon , Probe 和 Join 实现树的构建. 图 2 描述了消息交换构建融合树的过程.

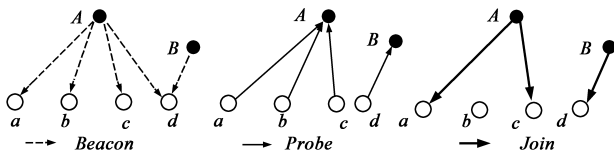


图 2 消息交换构建伸展树

从图 2 可以看出, A 选择节点 a 和 c 做为它们的孩子, 而 a 和 c 距离 A 比 a 和 b 或者 b 和 c 都远. 融合树构建算法 $\text{Form_AT}(p, p'')$ 的伪代码如下.

输入: 树的深度 p 和父节点选择概率.

输出: 深度为 p 的二叉树 T_n 且每一个节点分配唯一的 ID

Begin

1: For each level j from 0 to $p-1$

For each node i from 1 to 2^j

2: M_i is a node at level $j+1$

3: n_i is a node at level j

4: n_i sends Beacon packet containing n_i 's ID to M_i

Where distance between n_i and $M_i < r / * r$ 为通信半径 $*$

5: M_i chooses n_i as its parent with probability $> p''$ and the most path length of span

6: M_i sends Probe packet to n_i

7: n_i waits $NWAIT$ time (which is a sufficiently long fixed time period) to receive Probe packet from each M_i who selected n_i as parent

8: End

对于传感器网络来说, 在最大跨度路径上的节点具有较好的分散性, 减少了由于覆盖重叠而对网络感知能力所造成的影响, 因此这些分散性好的节点需要保持. 通过节点伸展树的定义, 确定了传感器网络中需要保持的节点集. 基于此节点集, 对于覆盖重叠节点也

可有效地提高网络整体的感知能力.

2.3 数据融合

基于伸展树的数据融合算法的主要思想是采用传输能够拟合较多的检测数据的模型 M 来代替传输节点的检测数据, 其目的是为了减少数据传输量, 从而节省传感器节点的能量. 因此需要考虑回传以参数表示的模型的代价和其可拟合的数据量之间的关系. 传输模型的代价越小其能够表示的数据越多, 节点就越节省能量. 由于节点检测值往往要受多个因素的影响, 我们期望用最小代价模型拟合最多的数据, 且计算简单, 而多元线性回归模型恰好符合这一目标.

融合树中的每个节点接收并存储由最近的非树节点周期性报告给它的的功能, 即 NT 节点负责感知而 AT 节点负责存储, 将存储在 AT 节点里的值看作是输入 $x-y$ 坐标的函数值, 此过程由三元组 (f, x, y) 描述, 这里 f 为位于 (x, y) 处节点的感知估计值, 由 AT 中存储在节点 i 的数据及其子节点发送的多元回归模型的系数执行多元线性函数回归生成. 设多元线性回归模型形式如下^[18]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \mu \quad (5)$$

其中: Y 为感知估计值, $X_j (j=1, 2, \dots, k)$ 为对感知估计值 Y 发生作用的影响因子, $\beta_j (j=0, 1, 2, \dots, k)$ 为 $k+1$ 个未知回归参数, μ 为随机误差项. 由于参数 $\beta_j (j=0, 1, 2, \dots, k)$ 都是未知的, 可以利用样本观测值 $(x_{1i}, x_{2i}, \dots, x_{ki}; Y_i)$ 对它们进行估计, 由此得到的参数估计值为 $\hat{\beta}_j (j=0, 1, 2, \dots, k)$, 用参数估计值替代回归模型的未知参数 $\beta_j (j=0, 1, 2, \dots, k)$, 则得多元线性样本回归方程:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_k x_{ki} \quad (6)$$

其中 $\hat{Y}_i (i=1, 2, \dots, n)$ 为 Y_i 的样本回归值. 则观测值 Y_i 与回归值 \hat{Y}_i 的残差 e_i 为:

$$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_k x_{ki}) \quad (7)$$

由最小二乘法可知 $\hat{\beta}_j (j=0, 1, 2, \dots, k)$ 应使全部观测值 Y_i 与回归值 \hat{Y}_i 的残差 e_i 的平方和最小, 即使

$$\begin{aligned} Q(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k) &= \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 \\ &= \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki})^2 \end{aligned} \quad (8)$$

取得最小值. 根据多元函数的极值原理, Q 分别对 $\hat{\beta}_j (j=0, 1, 2, \dots, k)$ 求一阶偏导, 并令其等于零.

$$\frac{\partial Q}{\partial \hat{\beta}_j} = 0, (j=1, 2, \dots, k) \quad (9)$$

将上述 $k+1$ 个方程化简得下列方程组

$$\begin{cases} n\hat{\beta}_0 + \hat{\beta}_1 \sum x_{1i} + \hat{\beta}_2 \sum x_{2i} + \dots + \hat{\beta}_k \sum x_{ki} = \sum Y_i \\ \hat{\beta}_0 \sum x_{1i} + \hat{\beta}_1 \sum x_{1i}^2 + \hat{\beta}_2 \sum x_{2i} x_{1i} + \dots + \hat{\beta}_k \sum x_{ki} x_{1i} = \sum x_{1i} Y_i \\ \dots \\ \hat{\beta}_0 \sum x_{ki} + \hat{\beta}_1 \sum x_{1i} x_{ki} + \hat{\beta}_2 \sum x_{2i} x_{ki} + \dots + \hat{\beta}_k \sum x_{ki}^2 = \sum x_{ki} Y_i \end{cases} \quad (10)$$

经化简得如下方程:

$$\mathbf{X}^T \mathbf{Y} = \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} \quad (11)$$

$$\hat{\boldsymbol{\beta}} = [\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k]^T \quad (12)$$

设 $R(X) = K + 1$, $\mathbf{X}^T \mathbf{X}$ 为 $(K + 1)$ 阶方阵, 则 $\mathbf{X}^T \mathbf{X}$ 满秩, 其逆矩阵存在, 所以 $\hat{\boldsymbol{\beta}}$ 的最小二乘估计向量为:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (13)$$

运用多项式回归, 得到如下方程.

$$\mathbf{X} = \begin{pmatrix} 1 & y_1 & y_1^2 & x_1 & x_1 y_1 & x_1 y_1^2 & x_1^2 & x_1^2 y_1 & x_1^2 y_1^2 \\ 1 & y_2 & y_2^2 & x_2 & x_2 y_2 & x_2 y_2^2 & x_2^2 & x_2^2 y_2 & x_2^2 y_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_n & y_n^2 & x_n & x_n y_n & x_n y_n^2 & x_n^2 & x_n^2 y_n & x_n^2 y_n^2 \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_8 \end{pmatrix} \quad \text{其中: } \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (14)$$

$$f(x, y) = \hat{\beta}_0 + \hat{\beta}_1 y + \hat{\beta}_2 y^2 + \hat{\beta}_3 x + \hat{\beta}_4 xy + \hat{\beta}_5 xy^2 + \hat{\beta}_6 x^2 + \hat{\beta}_7 x^2 y + \hat{\beta}_8 x^2 y^2 \quad (15)$$

从式(14)可知, 由给定的位置信息 (x, y) 计算 $\hat{\boldsymbol{\beta}}$, 通过式(15)获得的值 $f(x, y)$ 是 (x, y) 节点处的感知估计值. 由于 $\mathbf{X}^T \mathbf{X}$ 是 $k + 1$ 阶满秩. 换言之, $n \gg k + 1$ 且 \mathbf{X} 不能表示为任何其它列集合的加权线性组合. 本文提出的数据融合算法按宽度优先输入, 每个树节点由式(14)生成系数并把这个系数集发送给它的父节点, 每一层的节点从其孩子获得的系数更新感知值, 这些数据结合节点本身读入的检测值计算新的系数集, 然后向更高一层传递. 在这个过程中, 识别区域内产生的事件属性值是问题的关键, 因为它们直接影响到融合值的准确性, 可通过坐标 $\{x_{\min}, y_{\min}, x_{\max}, y_{\max}\}$ 的上下界识别该区域, 这里, 最大值和最小值取自子树里当前父节点之下的所有传感节点. 如图 3, 设 a 为当前的融合节点, 这个区域范围通过传感节点的最小和最大坐标传递给 a 之下的子树来界定. 这样 a 从它的孩子那里得到这个区域的坐标边界. 通过基于伸展树的构建和上面所描述的回归过程, 每隔指定时间间隔回答查询, 如“SELECT temperature FROM sensors WHERE location = (x, y) ”, 或“目标范围内最高温度”之类问题. 当 Sink 需要知道在 (x, y) 这个位置点的数据时, 它就向根节点发送这个查询, 查询由 AT 向下传播直到该节点所在的子区域, 在该子区域内, 由最近的感知节点报告检测数据到 (x, y) 处的树节点, 执行数据融合过程. 数据融合算法 SPAT $(p, n_s) / * n_s$ 为感知节点向树节点报告的平均数 $*/$

multivariate polynomial regression is performed on each datas and the coefficients are stored in the each of the arrays $\beta_0, \beta_1, \dots, \beta_8$ each of size N

end for

3: Initialize level to 2^p

while p is greater than 0

sum = level + 2^{p-1}

k = level

while $k < \text{sum}$

4: for each of the non leaf nodes k of the tree, k computes random $x - y$ points for each of its 2 children i and $(i + l)$ where (x_{\min}, y_{\min}) and (x_{\max}, y_{\max}) are the coordinates of the leftmost and down most node and right most and top most node respectively reporting to the node i and $i + l$.

endfor

5: Using $(\beta_{i0}, \beta_{i1}, \dots, \beta_{i8})$ and $(\beta_{(i+1)0}, \beta_{(i+1)1}, \dots, \beta_{(i+1)8})$ new attribute values are calculated and appended to node k ' data at node k then calls the regression function to calculate $(\beta_{k0}, \dots, \beta_{k8})$ and passes it to its parent.

6: endwhile

7: Level = sum

8: Endwhile

9: End

相对于传输原始数据到 Sink 节点的能量消耗和时延, 通过 SPAT 融合过程报告数据更为有效.

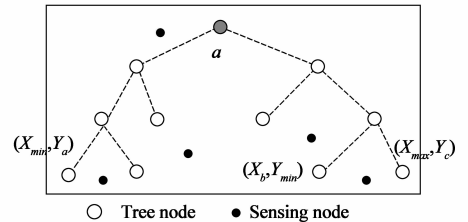


图3 融合过程中边界区域的计算

3 事件区域检测

3.1 事件区域检测读数的判定

本文对正常与错误读数 $\sigma(i, t)$ 的判定条件如下.

正常读数: 通常情况下, 检测读数 $\sigma(i, t)$ 在 $[C_{\min}, C_{\max}]$ 内且服从正态分布 $\varphi(i)$, 若存在 $\tau_1 > 0$, 使得检测读数满足 $|\sigma(i, t) - E(i)| \geq \tau_1$, 说明传感器读数超出了正常范围. 满足该条件的节点越多, 事件发生的可能性就越大.

错误读数: 对于传感器 i , $\varphi(i)$ 满足以下三个条件中的任一个, 该传感器可能为故障传感器, 其检测产生的读数为错误读数.

(1) 若存在 τ_1, τ_{th} 且 $\tau_{th} > \tau_1 > 0$, 对于 $\forall t, |\sigma(i, t) - E(i)| \geq \tau_{th}$, 说明传感器的读数超出了正常范围, 它很可能是有故障的. 如果传感器频繁或者持续报告这样的读数, 那么认为这个传感器是有故障的.

- 1: Begin
- 2: For each of leaf node i of the tree
node i ' data is read

4 仿真结果与讨论

本文采用离散事件仿真平台 NS2 进行模拟实验,采用无碰撞冲突的 MAC 协议,选择温度属性评估所提出方案的有效性,表 1 定义了所使用的参变量.设区域 A 中节点总数为 D ,则节点密度 $\rho = D/A$, A_s 是包含单一融合树 T_c 的子区域,因此在子区域里节点的平均数 U 由 A_s 确定,对于二叉树,节点总数为 $t = 2^{(p+1)} - 1$,此外, n_s 由前面设定,即感知节点向树节点报告的平均数,子区域里节点数 U 的上界为^[15]: $U = n_s \times t + t$ 或 $t = U/(n_s + 1)$,利用式(1)替换 t ,得 AT 深度的一个最优解:

$$p = \ln\left(\frac{U}{n_s + 1} + 1\right) - 1 \quad (16)$$

表 1 仿真实验参数设置

Parameter	Variation
区域面积 A	800 × 800
通信半径 R	40m
节点总数 D	1630
节点密度 A/D	0.0025
自事件区域 A_s	400 × 400
父节点选择概率 P^u	0.33
树的深度 p	4
报告的平均数 n_s	12

4.1 性能评估测度

在 800×800 平方单元区域 R 内,随机部署 250 个节点,温度变化 $30 \sim 35^\circ\text{C}$ 是正常现象,温度变化在 $39 \sim 49^\circ\text{C}$ 时为异常,可能是该区域发生了火灾.树的深度设置为 4.在无事件发生时,检测的温度值服从正态分布 $\varphi(i)$. C_{\min} 、 C_{\max} 和 $E(i)$ 分别为 30、35 和 32.5, $\varphi(i)$ 的变化是 1.2.单个事件 $P_{\text{event}1}$ 在中心发生,事件区域称为 R_{event} , $E(d_i) - C_{\min} = C_{\max} - E(d_i) = 2.5$, τ_1 为 2.5, $E(\sigma(i, t)) - C_{\min} = C_{\max} - E(\sigma(i, t)) = 2.5$ 且 $\tau_2 = \tau_{th} = 4.5$.当事件发生时,读取的数据通过 NT 节点传输到最近的树节点,通过 10 轮实验评估以下性能指标.前提是假设查询已经广播到融合树的所有叶节点.

(1) 错误率 (Percentage error): AT 中根节点获得的读数与实际传感器读数的近似程度.由事件区域检测的读数与实际读数产生的偏差 E 表示, $E = \left(\sum_{k=1}^{n_b} \frac{|z_k - \bar{z}_k|}{z_k} \times 100\right) / n_b \leq \epsilon_{th}$,其中 $\epsilon_{th} = 10\%$ 是指定的错误阈值, n_b 是在边界传感器的数量, \bar{z} 为近似读数, z 为实际读数.

(2) 事件检测延迟 (Event recognition delay): 事件发生的时间与在根节点计算的近似时间之间的间隔.

4.2 仿真结果

4.2.1 错误率

当通信范围变化时,对 STERD 完成事件检测的准确性进行分析,如图 6 所示,两事件 STERD 的误差率不足 10%,采用多事件 STERD 进行事件检测(误差范围为 6.8~8.9)比单 STERD 进行事件检测(误差范围为 6.4~7.6)高出 6%,单事件检测出现较小的误差范围,是由于近似边界可以通过最后生成的 f_{event} 较好定位,而在多事件 STERD 方法中是无法实现的.由图 6 可以看出,被检测事件的错误级别随通信范围的增长呈不显著变化,而图 5 中,随着通信范围的增加,在区域 R_{event} 里有更多的传感器读数被传送到树,错误级别却随通信范围的增长急剧增加,在 STERD 中,不同的读数与不同的事件相对应,树节点会生成不同的多项式,总体误差率是独立于各个传感器的通信范围的.从图 5 还可以看出,随着 NT 节点的增加,错误率缓慢增加,树节点、 NT 节点位于事件边界的可能性也会增加,回归多项式的准确性也会增加,进而覆盖到相对大的范围和提供更加准确的传感参数,错误率被控制在一个合适的范围内.

4.2.2 检测延迟

图 6 表明,事件检测延迟随节点密度的增加几乎保持不变.导致事件检测延迟的主要因素有 3 个方面:事件识别延迟、多项式计算延迟和事件报告的传输延迟(称为通信延迟).从图 6 可以看出,事件识别和多项式计算延迟远远小于通信延迟,通信延迟占总延迟的近 77%.一旦融合树确定,用于通信的延迟保持恒定并且独立于 NT 节点的密度.当节点密度增加,围绕每棵树的 NT 节点数目也随之增加.实验结果表明,当节点数

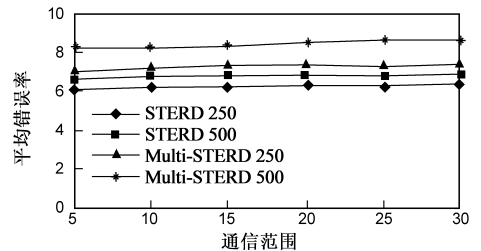


图5 通信范围、节点密度与错误率的关系

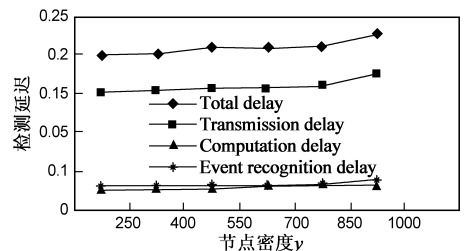


图6 事件检测、节点密度对延迟的影响

目 n_m 增加时,事件识别、计算负载和多项式构造也维持恒定,这是由于节点数目的增加对 n_s 的影响不大. STERD 之所以能够降低事件识别和事件报告的复杂度是由基于以下原因:首先,在 STERD 中,事件识别和基于多项式的数据融合不含有任何复杂计算;其次,基于树的网络结构对事件识别进行了局部处理;再者,当网络规模增加时,树节点的数目也相应增加,因此 STERD 是可扩展的.

5 结束语

本文研究了无线传感器网络事件区域的容错检测.首先,通过构建融合树提出了一种新的数据融合算法,该算法在缺少传感节点的位置,利用被检测事件的时空相关性,也能获得事件检测的属性值,其错误率控制在可接受的范围内.另外,在此基础上,提出了一种基于融合树的事件区域检测容错算法,该算法可以对多个事件进行检测,识别发生在边界区域的事件和有故障的传感器节点,快速把检测读数传给基站,数据压缩比基本上是恒定的,减少了数据传输的能量消耗和通信延迟.

参考文献:

[1] 李建中, 李金宝, 石胜飞. 传感器网络及其数据管理的概念问题与进展[J]. 软件学报, 2003, 14(10): 1717 - 1727.
Li Jian-zhong, Li Jin-bao, Shi Sheng-fei. Concepts, Issues and Advance of Sensor Networks and Data Management of Sensor Networks[J]. Journal of Software, 2003, 14(10): 1717 - 1727. (in Chinese)

[2] 蔚赵春, 周水庚, 关倩红. 无线传感器网络中数据存储与访问研究进展[J]. 电子学报, 2008, 36(10): 2002 - 2010.
Yu Zhao-chun, Zhou Shui-geng, Guan Ji-hong. Data storage and access in wireless sensor Networks: a survey [J]. Acta Electronica Sinica, 2008, 36(10): 2002 - 2010. (in Chinese)

[3] 朱艺华, 沈丹丹, 吴万登, 沈振伟, 汤一平. 无线传感器网络优化生存时间的动态路由算法[J]. 电子学报, 2009, 27(5): 1041 - 1045.
Zhu Yi-hua, Shen Dan-dan, Wu Wan-deng. Dynamic Routing Algorithms Optimizing Lifetime of Wireless Sensor Networks [J]. Acta Electronica Sinica, 2009, 27(5): 1041 - 1045. (in Chinese)

[4] Krishnamachari B, Iyengar S S. Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks[J]. IEEE Trans. on Computers, 2004, 53(3): 241 - 250.

[5] Chen Q, Lam K Y, Fan P. Comments on, distributed Bayesian algorithms for fault tolerant event region detection in wireless sensor networks[J]. IEEE Transactions on Computers, 2005, 54(9): 1182 - 1183.

[6] Luo X, Dong M, Huang Y. On distributed fault-tolerant detection in wireless sensor networks[J]. IEEE Trans on Computers, 2006, 55(1): 58 - 70.

[7] Koushanfar F, Potkonjak M, Sangiovanni-Vincentelli A. On-Line fault detection of sensor measurements[A]. In Proc of the IEEE Sensors[C]. NY USA: IEEE Computer Society, 2003, 2: 974 - 979.

[8] Ding M, Chen D et al. Localized fault tolerant event boundary detection in sensor networks[A]. In Proceedings of the Annual IEEE Conference on Computer Communications (INFOCOM), (C). Miami, 2005, 2: 902 - 913.

[9] Koushanfar F, Potkonjak M, Sangiovanni-Vincentelli A. Fault-Tolerance in sensor networks[M]. Handbook of Sensor Networks, CRC Press, 2004.

[10] Kui Ren, Kai Zeng, Wenjing Lou. Secure and fault-tolerant event boundary detection in wireless sensor Networks[J]. IEEE Transactions on Wireless Communications, 2008, 7(1): 354 - 363.

[11] Chen J, Kher S, Somani A. Distributed fault detection of wireless sensor networks[A]. In Proc of the 2006 Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks (DIWANS 2006)[C]. NY USA: ACM 2006. 65 - 72.

[12] Koushanfar F, Potkonjak M, Sangiovanni-Vincentelli A. Error models for light sensors by statistical analysis of raw sensor measurements[A]. In Proc of the IEEE Sensors[C]. 2004, 3, 1472 - 1475.

[13] R Nowak, U Mitra. Boundary estimation in sensor networks: theory and methods[A]. In Proc 2nd International Workshop on IPSN[C]. Berlin: Springer, 2003. 2634, 86 - 95.

[14] K K Chintalapudi, R Govindan, Localized edge detection in sensor fields[A]. In Proc IEEE Workshop on Sensor Network Protocols and Applications[C]. Los Angeles, Elsevier, 2003. 59 - 70.

[15] C Guestrin, P Bodi, R Thibau, et al, Distributed Regression: an Efficient Framework for Modeling Sensor Network Data [M.]. IPSN. New York, ACM Press, 2004.

[16] A Sharaf, J Beaver, A Labrinidis, et al. Balancing energy efficiency and quality of aggregate data in sensor networks[J]. The VLDB Journal, 2004, 13(4): 384 - 403.

[17] Aurenhammerf, KleinR. Handbook of Computational Geometry [M]. Amsterdam, Netherlands: North-Holland, 2000. 201 - 290.

[18] Michael Berthold, Intelligent Data Analysis [M]. Springer Press, 2nd Edition, 2007.

[19] Vuran MC, Akan OB, Akyildiz IF. Spatio-Temporal correlation: theory and application for wireless sensor networks[J]. Computer Networks, 2004, 45: 245 - 259.

作者简介:



张书奎 男,1966 年生于内蒙乌前旗,苏州大学计算机科学与技术学院副教授,博士,主要研究方向为无线网络、信息安全、分布式计算等.
E-mail: zhangsk2000@163.com



樊建席 男,1965 年生于山东青岛,苏州大学计算机科学与技术学院教授,博士生导师,主要研究方向为并行与分布式系统、网络互连结构设计及算法分析等.



崔志明 男,1961 年生于江苏苏州,苏州大学计算机科学与技术学院教授,博士生导师,主要研究方向为智能信息处理、计算机网络.作为项目负责人主持的国家自然科学基金、教育部重点科研项目等共 18 项;发表的论文 100 余篇,其中被 SCI、EI 收录 28 篇.



张文哲 男,1979 生于山西山阴,博士,苏州大学计算机科学与技术学院讲师,主要研究方向为无线传感器网络与普适计算.曾在软件学报、MSN、ICISS 等国内外重要期刊和会议上发表论文 10 余篇.